

## TESS: A Scalable Temporally and Spatially Local Learning Rule for Spiking Neural Networks

Apolinario , M.P.E.; Roy, K.; Frenkel, C.

**DOI**

[10.1109/IJCNN64981.2025.11227652](https://doi.org/10.1109/IJCNN64981.2025.11227652)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN)

**Citation (APA)**

Apolinario , M. P. E., Roy, K., & Frenkel, C. (2025). TESS: A Scalable Temporally and Spatially Local Learning Rule for Spiking Neural Networks. In *Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN)* (Proceedings of the International Joint Conference on Neural Networks). IEEE. <https://doi.org/10.1109/IJCNN64981.2025.11227652>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

**Green Open Access added to [TU Delft Institutional Repository](#)  
as part of the Taverne amendment.**

More information about this copyright law amendment  
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:  
the publisher is the copyright holder of this work and the  
author uses the Dutch legislation to make this work public.

# TESS: A Scalable Temporally and Spatially Local Learning Rule for Spiking Neural Networks

Marco P. E. Apolinario<sup>1,2</sup>, Kaushik Roy<sup>1</sup>, Charlotte Frenkel<sup>2</sup>

<sup>1</sup>*School of Electrical and Computer Engineering, Purdue University, Indiana, USA*

<sup>2</sup>*Microelectronics Department, Delft University of Technology, Delft, NL*

mapolina@purdue.edu, kaushik@purdue.edu, c.frenkel@tudelft.nl

**Abstract**—The demand for low-power inference and training of deep neural networks (DNNs) on edge devices has intensified the need for algorithms that are both scalable and energy-efficient. While spiking neural networks (SNNs) allow for efficient inference by processing complex spatio-temporal dynamics in an event-driven fashion, training them on resource-constrained devices remains challenging due to the high computational and memory demands of conventional error backpropagation (BP)-based approaches. In this work, we draw inspiration from biological mechanisms such as eligibility traces, spike-timing-dependent plasticity, and neural activity synchronization to introduce TESS, a temporally and spatially local learning rule for training SNNs. Our approach addresses both temporal and spatial credit assignments by relying solely on locally available signals within each neuron, thereby allowing computational and memory overheads to scale linearly with the number of neurons, independently of the number of time steps. Despite relying on local mechanisms, we demonstrate performance comparable to the backpropagation through time (BPTT) algorithm, within  $\sim 1.4$  accuracy points on challenging computer vision scenarios relevant at the edge, such as the IBM DVS Gesture dataset, CIFAR10-DVS, and temporal versions of CIFAR10, and CIFAR100. Being able to produce comparable performance to BPTT while keeping low time and memory complexity, TESS enables efficient and scalable on-device learning at the edge.<sup>1</sup>

**Index Terms**—Spiking Neural Networks, Local Learning Rule, On-device Learning

## I. INTRODUCTION

With the increasing ubiquity of low-power electronic devices and the rapid advances in artificial intelligence, particularly in deep neural networks (DNNs), there has been a growing interest in bringing intelligence to the edge [1]–[3]. The conventional approach, known as offline training, involves training DNNs in the cloud, where computational resources are abundant, and subsequently deploying the trained models on edge devices. However, certain use cases, such as those involving privacy concerns or the need for real-time model adaptation, render the offline approach unsuitable. In these scenarios, an on-device learning paradigm is essential. This approach requires the development of energy-efficient models and DNN learning rules that operate within the constraints of edge devices [4], [5].

This work was supported in part by the Center for Co-design of Cognitive Systems (CoCoSys), one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program, in part by the Department of Energy (DoE), and in part by the NSF AccelNet NeuroPAC Fellowship Program.

<sup>1</sup>Code available at <https://github.com/mapolinario94/TESS>

In recent years, biologically plausible models such as spiking neural networks (SNNs) have got attention as energy-efficient alternatives to conventional DNN, owing to their unique spatio-temporal processing capabilities, event-driven nature, and binary spiking activations [6]–[8]. While these features make SNNs promising candidates for enabling energy-efficient inference at the edge, new solutions for solving both the spatial and temporal credit assignment problems are needed for resource-constrained scenarios. For example, in an SNN with  $L$  layers and  $n$  neurons per layer, the backpropagation through time (BPTT) algorithm (Fig. 1a) exhibits time and memory complexities of  $O(TLn^2)$  and  $O(TLn)$ , respectively, where  $T$  denotes the length of the input sequence. This dependency on  $T$  makes BPTT impractical for on-device learning on low-power edge devices [8]–[10].

To address these limitations, several alternatives have been proposed to achieve local credit assignment. For *spatial credit assignment*, methods such as feedback alignment (FA) and direct feedback alignment (DFA) employ random matrices to propagate error signals or directly project errors to individual layers, thereby reducing layer dependencies [11]–[13]. Similarly, the direct random target projection (DRTP) method [14] projects targets generated from classification labels instead of output errors, allowing for independent and forward layer updates. Other biologically plausible approaches replace the backward pass in BP with an additional forward pass [15], [16]. Although these methods show promise, they often suffer from slow convergence and scalability issues when applied to deep networks. For instance, DFA experiences a  $\sim 16\%$  accuracy drop compared to BP on CIFAR10 in a five-layer model [17], while the method proposed in [15] does not scale beyond two-layer models. Recently, [17] introduced a local learning rule inspired by neural activity synchronization (LLS), which uses fixed periodic basis vectors for layer-wise training, demonstrating performance comparable to BP in fairly large datasets, including CIFAR100, and Tiny ImageNet.

For *temporal credit assignment* in SNNs, methods inspired by three-factor learning rules leverage eligibility traces [18] to preserve temporal information in a way that is biologically plausible. For example, [9] introduced e-prop, a method that uses eligibility traces to address temporal credit assignment in SNNs with explicit recurrent connections, while relying on BP or DFA for spatial credit assignment. Other approaches, such as [19], [20], combine eligibility traces with DRTP

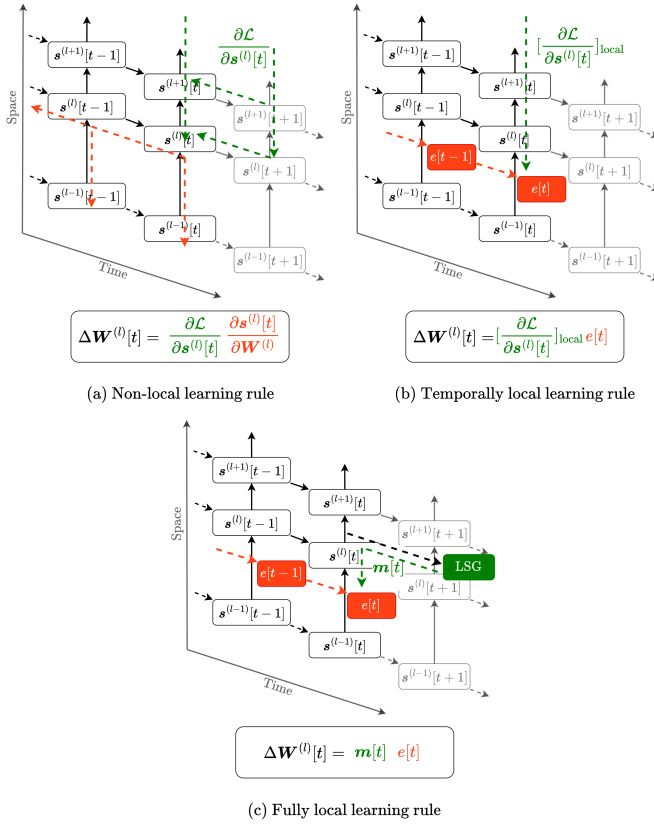


Fig. 1. Comparison of three learning rule strategies for training a recurrent model with state variable  $s[t]$ . (a) Non-local learning method (e.g., BPTT): Both spatial and temporal credit assignment problems are solved by propagating errors through time and space (layers). (b) Temporal local method: Temporal credit assignment is addressed using **eligibility traces** ( $e[t]$ ), which are auxiliary variables that track the history of neural activity. These traces are modulated by a **learning signal** ( $[\frac{\partial \mathcal{L}}{\partial s^{(l)}[t]}]_{\text{local}}$ ), which propagates errors across layers but not through time. (c) Fully local method (e.g., our proposed method TESS): In addition to eligibility traces, the **learning signal** ( $m[t]$ ) is generated locally, addressing both spatial and temporal credit assignment entirely within the local context.

for spatial credit assignment, achieving spatial and temporal locality. However, these methods have proven effective only for shallow models (2 – 3 layers) and scale poorly to deeper architectures due to their high memory cost of  $O(Ln^2)$ , which, while independent of the number of time steps, scales with the number of synapses ( $n^2$ ). Similarly, [21] uses eligibility traces in combination with auxiliary spiking classifiers, significantly increasing the number of additional trainable parameters. An alternative method, the spike-timing-dependent plasticity (STDP)-inspired temporal local learning rule (S-TLLR) [22], achieves performance comparable to BPTT with 5 – 50× less memory cost and 1.3 – 6.6× less multiply-accumulate (MAC) operations. Specifically, S-TLLR memory requirements scale linearly with the number of neurons instead of synapses and independent of the number of time steps, that is  $O(Ln)$ . However, S-TLLR still relies on backpropagating errors across layers for spatial credit assignment, as shown for temporally

local learning rules in Fig. 1b.

To overcome these limitations and enable efficient on-device learning, we propose TESS, a novel scalable temporally and spatially local learning rule for training SNNs (Fig. 1c). Unlike prior works, which either suffer from scalability issues (e.g., [9], [19], [20]) or rely on global error propagation across layers (e.g., [22], [23]), TESS fundamentally addresses the bottlenecks of memory and computational overhead in a way that is both temporally and spatially local. Specifically, TESS addresses the temporal credit assignment problem using eligibility traces with memory complexity that scales linearly with the number of neurons, drawing inspiration from [22]. For the spatial credit assignment problem, TESS employs a mechanism that synchronizes the activity of neurons within each layer by modulating eligibility traces with a locally generated learning signal, derived from fixed basis vectors inspired by [17]. This entirely local approach eliminates the need for backpropagation across layers, unlocking scalability to deeper architectures and larger datasets while maintaining memory and computational complexity that are independent of the number of time steps, that is  $O(Ln)$  and  $O(LCn)$ , respectively.

Crucially, TESS marks a significant advancement over prior works by introducing, for the first time, a low-complexity fully local training method for SNNs that achieves performance comparable to BPTT. On challenging datasets such as CIFAR10-DVS, TESS incurs only a  $\sim 1.4\%$  accuracy drop relative to BPTT, while on other datasets, including IBM DVS Gesture, and temporal version of CIFAR10, and CIFAR100, TESS matches the performance of BPTT. Notably, this is achieved while drastically reducing both computational and memory requirements, with 205 – 661× fewer MACs and 3 – 10× lower memory usage. This capability is achieved through the integration of biologically plausible mechanisms like eligibility traces, STDP, and neural activity synchronization, all of which rely solely on locally available signals within each neuron.

The main contributions of this work are summarized as follows:

- We propose TESS, a novel scalable learning rule for SNNs that integrates biologically inspired mechanisms such as eligibility traces, STDP, and neural activity synchronization. TESS operates in a fully local manner, relying only on locally available signals, making it well-suited for energy-efficient on-device learning.
- TESS achieves linear memory complexity,  $O(Ln)$ , and computational complexity,  $O(LCn)$ , enabling the efficient training of deeper architectures, such as VGG-9, on resource-constrained edge devices.
- TESS delivers performance comparable to BPTT on vision benchmarks, matching its accuracy on IBM DVS Gesture, CIFAR10, and CIFAR100, while incurring only a  $\sim 1.4\%$  accuracy drop on CIFAR10-DVS. This is achieved with significantly lower resource requirements.

## II. BACKGROUND

In this section, we introduce the spiking neuron models adopted here, the mathematical notation, gradient-based optimization approaches for SNNs, and three factor learning rules.

### A. LIF model

We adopt the leaky integrate-and-fire (LIF) neuron model to implement spiking behavior. The discrete LIF neuron model is mathematically described as follows:

$$\mathbf{u}_i^{(l)}[t] = \gamma(\mathbf{u}_i^{(l)}[t-1] - v_{\text{th}}\mathbf{o}_i^{(l)}[t-1]) + \sum_j \mathbf{W}_{ij}^{(l)}\mathbf{o}_j^{(l-1)}[t], \quad (1)$$

$$\mathbf{o}_i^{(l)}[t] = \Theta(\mathbf{u}_i^{(l)}[t] - v_{\text{th}}), \quad (2)$$

where,  $\mathbf{u}_i^{(l)}[t]$  represents the membrane potential of the  $i$ -th neuron in layer  $l$  at the time step  $t$ , and  $\mathbf{W}_{ij}^{(l)}$  is the strength of the synaptic connection between the  $i$ -th post-synaptic neuron in layer  $l$  and the  $j$ -th pre-synaptic neuron in layer  $l-1$ . The parameter  $\gamma$  is the leaky factor, producing an exponential decay of the membrane potential over time. The threshold voltage is denoted by  $v_{\text{th}}$ , and  $\Theta$  represents the Heaviside function ( $\Theta(x) = 1$  if  $x > 0$  and 0 otherwise). When  $\mathbf{u}_i^{(l)}[t]$  reaches  $v_{\text{th}}$ , the neuron fires, producing a binary spike output  $\mathbf{o}_i^{(l)}[t] = 1$ . This firing triggers a subtractive reset mechanism, represented by the reset signal  $v_{\text{th}}\mathbf{o}_i^{(l)}[t]$ , which reduces  $\mathbf{u}_i^{(l)}[t]$  by  $v_{\text{th}}$ .

### B. Gradient-based optimization for SNNs

We now describe how gradient-based optimization is applied to SNNs.

Given a dataset  $\mathbb{D} = \{(\mathbf{x}, \mathbf{y}^*)_{i=1}^N\}$ , where  $N$  is the number of samples,  $\mathbf{x}$  represents the input data, and  $\mathbf{y}^*$  denotes the corresponding labels, and an SNN model with parameters  $\mathbb{W} = \{\mathbf{W}^{(l)}\}_{l=1}^L$ , where  $L$  is the number of layers, the optimization goal is to minimize a loss function  $\mathcal{L}$ ,

$$\mathbb{W} := \arg \min_{\mathbb{W}} \mathcal{L}(\mathbb{D}; \mathbb{W}).$$

This minimization is solved using (stochastic) gradient descent, where the parameters are iteratively updated as:

$$\mathbf{W}^{(l)} := \mathbf{W}^{(l)} - \eta \frac{d\mathcal{L}}{d\mathbf{W}^{(l)}},$$

where  $\eta$  is the learning rate and  $\frac{d\mathcal{L}}{d\mathbf{W}^{(l)}}$  represents the gradient of the loss function with respect to the parameters of the  $l$ -th layer. The gradients are computed using the BPTT algorithm, which applies the chain rule over both space (i.e. layers) and time:

$$\frac{d\mathcal{L}}{d\mathbf{W}^{(l)}} = \sum_{t=1}^T \frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t]} \frac{\partial \mathbf{u}^{(l)}[t]}{\partial \mathbf{W}^{(l)}},$$

where  $T$  is the total number of time-steps of the input sequence  $\mathbf{x}$ . Due to the recurrent nature of SNNs,  $\frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t]}$  depends on the entire history of the model:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t]} = \frac{\partial \mathcal{L}}{\partial \mathbf{o}^{(l)}[t]} \frac{\partial \mathbf{o}^{(l)}[t]}{\partial \mathbf{u}^{(l)}[t]} + \frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t+1]} \frac{\partial \mathbf{u}^{(l)}[t+1]}{\partial \mathbf{u}^{(l)}[t]},$$

where the term  $\frac{\partial \mathcal{L}}{\partial \mathbf{o}^{(l)}[t]}$  requires information from all subsequent layers ( $L-l$ ), while  $\frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t+1]} \frac{\partial \mathbf{u}^{(l)}[t+1]}{\partial \mathbf{u}^{(l)}[t]}$  depends on the full temporal history. Thus, BPTT is neither spatially nor temporally local and incurs high computational costs.

### C. Three-factor learning rules

Three-factor learning rules [18] represent a biologically plausible framework for synaptic plasticity, where synapse updates are determined by the interaction of three key factors: pre-synaptic activity, post-synaptic activity, and a top-down learning signal.

The core idea of three-factor learning rules is the concept of an eligibility trace ( $\mathbf{e}_{ij}$ ), which is a transient variable that encodes synaptic changes driven by pre- and post-synaptic activity. This trace persists over time, allowing updates to occur when a delayed top-down learning signal arrives. The eligibility trace is typically modeled as a function of the pre- and post-synaptic activity, decaying over time according to the following recurrent formulation:

$$\mathbf{e}_{ij}^{(l)}[t] = \beta \mathbf{e}_{ij}^{(l)}[t-1] + f(\mathbf{o}_i^{(l)}[t])g(\mathbf{o}_j^{(l-1)}[t]), \quad (3)$$

where  $\beta$  is an exponential decay factor,  $f(\mathbf{o}_i^{(l)}[t])$  and  $g(\mathbf{o}_j^{(l-1)}[t])$  are element-wise functions of the post-synaptic activity  $\mathbf{o}_i^{(l)}[t]$  and pre-synaptic activity  $\mathbf{o}_j^{(l-1)}[t]$ , respectively. This formulation ensures that synapses are “eligible” for updates only when neuronal activity meets certain conditions.

The actual synaptic update is computed by modulating the eligibility trace with a top-down learning signal  $\mathbf{m}_i[t]$ , which represents error or reward information. The weight update rule can be expressed as:

$$\Delta \mathbf{W}_{ij} = \sum_t \mathbf{m}_i[t] \mathbf{e}_{ij}^{(l)}[t], \quad (4)$$

where the learning signal  $\mathbf{m}_i[t]$  is typically derived from task-relevant objectives, such as the gradient of a loss function or a reward signal. This modulation ensures that synaptic updates are oriented towards minimizing a particular objective.

Three-factor learning rules have demonstrated effectiveness in training artificial SNNs, as evidenced by [9], [10], [20], [22]–[24]. Notably, they offer a biologically plausible approximation of BPTT under specific conditions [9], [25]. Despite their promise, previous works leveraging eligibility traces for temporal credit assignment still rely on global error propagation across layers, Fig. 1b, for achieving performance comparable to BPTT [10], [22], [23]. Which results in a time complexity of  $O(Ln^2)$ .

## III. PROPOSED METHOD - A SCALABLE FULLY LOCAL LEARNING RULE

To address the challenges of temporal and spatial credit assignment for SNNs, we propose TESS, a scalable temporally and spatially local learning rule that is biologically inspired. It is designed as a three-factor learning rule that operates efficiently with low computational and memory overhead, making it suitable for resource-constrained edge devices. Specifically,

TESS is presented in Fig. 2 and relies on two key components to achieve temporal and spatial locality:

*a) Temporal Credit Assignment with Eligibility Traces:*

As discussed in Section II-C, eligibility traces are transient variables driven by changes in pre- and post-synaptic activity. These traces encode the temporal history of synaptic connections, identifying synapses as candidates for updates and thereby addressing the temporal credit assignment problem by tracking neuronal activity.

However, eligibility traces as formulated in (3) incur a memory complexity of  $O(n^2)$ , which is prohibitive for deep SNN models. To overcome this, and in line with prior works [4], [22], [23], we restrict the formulation to instantaneous eligibility traces by setting  $\beta = 0$  in (3). This modification reduces the memory complexity to  $O(n)$  by independently tracking pre- and post-synaptic activity traces.

In TESS, we utilize two eligibility traces: one based on pre-synaptic activity (shown in red in Fig. 2) and one based on post-synaptic activity (shown in blue in Fig. 2). These two components mimic STDP mechanisms, capturing causal (red signals) and non-causal (blue signals) dependencies between pre- and post-synaptic activity.

We first consider the eligibility trace with causal information. Using (3), the function  $f(\cdot)$  is defined as a secondary activation function  $\Psi(\cdot)$  applied to the membrane potential  $\mathbf{u}^{(l)}[t]$ . This serves a role analogous to surrogate gradients in gradient-based approaches [26]. The function  $g(\cdot)$  is a low-pass filter applied to the input spikes:  $\sum_{t'=0}^t \lambda_{\text{pre}}^{t-t'} \mathbf{o}^{(l-1)}[t']$ , where  $\lambda_{\text{pre}}$  is an exponential decay factor, representing the trace of pre-synaptic activity. To compute this trace in a forward-in-time manner, we introduce a recurrent variable  $\mathbf{q}^{(l)}[t]$ , defined as:

$$\mathbf{q}^{(l)}[t] = \lambda_{\text{pre}} \mathbf{q}^{(l)}[t-1] + \mathbf{o}^{(l-1)}[t], \quad (5)$$

which allows the causal eligibility trace to be expressed as:

$$\mathbf{e}_{\text{pre}}^{(l)}[t] = \alpha_{\text{pre}} \Psi(\mathbf{u}^{(l)}[t]) \otimes \mathbf{q}^{(l)}[t], \quad (6)$$

where  $\alpha_{\text{pre}}$  controls the amplitude of the eligibility trace. For all experiments,  $\alpha_{\text{pre}}$  is set to 1.

For the second eligibility trace,  $\mathbf{e}_{\text{post}}^{(l)}[t]$ , we use a low-pass filter over the activations of the membrane potential  $\Psi(\mathbf{u}^{(l)}[t])$ :  $\sum_{t'=0}^{t-1} \lambda_{\text{post}}^{t-t'} \Psi(\mathbf{u}^{(l)}[t'])$ . This can also be expressed as a recurrent equation by introducing a new variable  $\mathbf{h}^{(l)}[t]$ :

$$\mathbf{h}^{(l)}[t] = \lambda_{\text{post}} \mathbf{h}^{(l)}[t-1] + \Psi(\mathbf{u}^{(l)}[t-1]), \quad (7)$$

and the non-causal eligibility trace is then given by:

$$\mathbf{e}_{\text{post}}^{(l)}[t] = \alpha_{\text{post}} \mathbf{h}^{(l)}[t] \otimes \mathbf{o}^{(l-1)}[t], \quad (8)$$

where  $\alpha_{\text{post}}$  determines the inclusion of non-causal terms, with  $\alpha_{\text{post}} = +1$  for positive inclusion,  $\alpha_{\text{post}} = -1$  for negative inclusion, and  $\alpha_{\text{post}} = 0$  for exclusion.

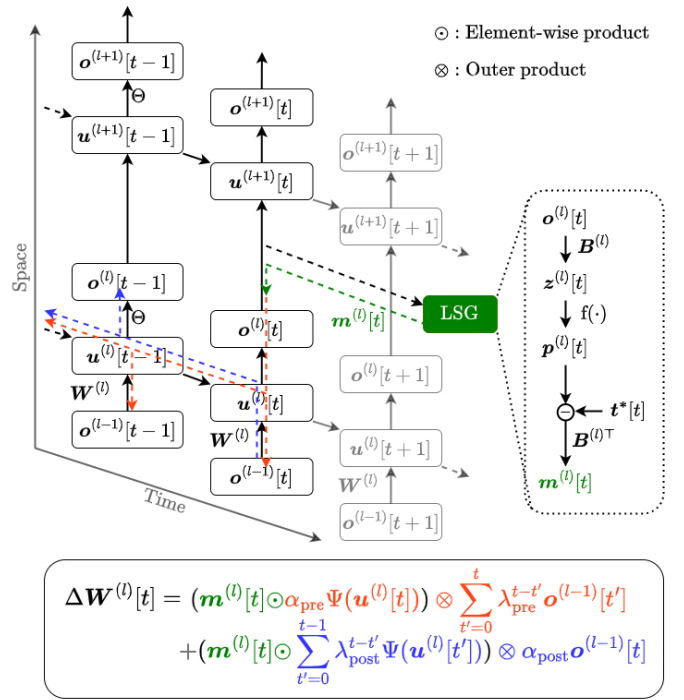


Fig. 2. Overview of TESS. The diagram illustrates an SNN model unrolled in time, where  $\mathbf{u}^{(l)}[t]$  denotes the membrane potential of neurons in the  $l$ -th layer at time step  $t$ , and  $\mathbf{o}^{(l)}[t]$  represents the corresponding output spikes. Signals involved in weight update computation are highlighted: red represents the eligibility trace based on causal relationships between inputs and outputs, blue represents the eligibility trace for non-causal relationships, and green represents the local learning signal  $\mathbf{m}^{(l)}[t]$  used to modulate the eligibility traces. The local learning signal is generated independently for each layer through a learning signal generation (LSG) process. The fixed binary matrix  $\mathbf{B}^{(l)}$  used in the LSG process features columns corresponding to square wave functions. While,  $f(\cdot)$  is a softmax function, and  $\mathbf{t}^*[t]$  represent the labels.

*b) Spatial Credit Assignment with Locally Generated Learning Signals:*

As discussed in Section II-C, while eligibility traces address the temporal credit assignment, synaptic updates require a top-down learning signal, denoted as  $\mathbf{m}^{(l)}[t]$ , to modulate the eligibility traces and solve the spatial credit assignment problem. Unlike prior approaches that rely on the global backpropagation of errors to compute this learning signal, TESS introduces a local mechanism for spatial credit assignment through a learning signal generation (LSG) process. This method, depicted in Fig. 2, enables the learning signals to be computed locally within each layer, avoiding the computational overhead of global error propagation and making the approach both scalable and hardware-friendly.

The LSG process begins by projecting the output spikes of each layer,  $\mathbf{o}^{(l)}[t]$ , into a  $C$ -dimensional task subspace using a projection matrix  $\mathbf{B}^{(l)}$ . This projection captures the relevant task information for the layer. Next, a function  $f(\cdot)$  is applied to the projected vector to compute its alignment with the target vector  $\mathbf{y}^*$ . The choice of  $f(\cdot)$  depends on the task: for classification problems,  $f(\cdot)$  is the softmax function, while for regression tasks, it is simply the identity function.

The difference between the projected vector and the target,  $f(\mathbf{B}^{(l)}\mathbf{o}^{(l)}[t]) - \mathbf{y}^*$ , serves as the error signal. This error signal is then projected back to the layer using the transpose of the projection matrix,  $\mathbf{B}^{(l)\top}$ , resulting in the modulatory learning signal,  $\mathbf{m}^{(l)}[t]$ , which is given by:

$$\mathbf{m}^{(l)}[t] = \mathbf{B}^{(l)\top} \left( f(\mathbf{B}^{(l)}\mathbf{o}^{(l)}[t]) - \mathbf{y}^* \right). \quad (9)$$

The design of the projection matrix  $\mathbf{B}^{(l)}$  is a critical component of the LSG process. In TESS,  $\mathbf{B}^{(l)}$  is defined as a fixed binary matrix, with each column corresponding to a square wave function. This design offers several advantages. The square wave functions help synchronize the activity of neurons within the same layer by assigning distinct spatial frequencies to different classes, ensuring that the task-related information is distributed effectively across the layer. Furthermore, the columns of  $\mathbf{B}^{(l)}$  are quasi-orthogonal, minimizing interference between the projections of different classes. The simplicity of square wave functions also makes them highly hardware-efficient, which is particularly advantageous for implementation in resource-constrained environments.

The locally generated learning signal in TESS eliminates the need for global backpropagation, significantly reducing computational complexity from  $O(n^2)$  to  $O(Cn)$ . The effectiveness of this design has been demonstrated in prior work [17], further validating its potential for real-world applications.

*c) Weight Updates:* The weight updates are computed based on the interaction of the learning signals with the eligibility traces. The updates for causal terms ( $\mathbf{W}_{\text{pre}}^{(l)}$ ) and non-causal terms ( $\mathbf{W}_{\text{post}}^{(l)}$ ) contributions are given by:

$$\Delta \mathbf{W}_{\text{pre}}^{(l)}[t] = \left( \mathbf{m}^{(l)}[t] \odot \alpha_{\text{pre}} \Psi(\mathbf{u}^{(l)}[t]) \right) \otimes \mathbf{q}^{(l)}[t], \quad (10)$$

$$\Delta \mathbf{W}_{\text{post}}^{(l)}[t] = \left( \mathbf{m}^{(l)}[t] \odot \alpha_{\text{post}} \mathbf{h}^{(l)}[t] \right) \otimes \mathbf{o}^{(l-1)}[t]. \quad (11)$$

The total weight update combines these two contributions:

$$\Delta \mathbf{W}^{(l)}[t] = \Delta \mathbf{W}_{\text{pre}}^{(l)}[t] + \Delta \mathbf{W}_{\text{post}}^{(l)}[t] \quad (12)$$

### A. Algorithm Implementation

The TESS algorithm operates iteratively, updating eligibility traces, computing learning signals, and adjusting weights for each time step. A pseudo-code implementation for layer  $l$  is provided in Algorithm 1.

### B. Computational and Memory Cost

In this subsection, we analyze the theoretical computational improvements of TESS in terms of multiply-accumulate (MAC) operations and memory requirements, comparing it to BPTT and S-TLLR. We build on the analysis presented in [22], which we expanded to include the effects of the spatial and temporal locality of TESS on computational and memory costs.

---

### Algorithm 1 TESS pseudo-code for layer $l$

---

**Input:**  $\mathbf{o}^{(l-1)}$  (input for layer  $l$ ),  $\mathbf{B}$  (fixed binary matrix),  $\beta$  (threshold),  $\eta$  (learning rate),  $t_l$  (time step to start generating the learning signal)

**Output:**  $\mathbf{W}^{(l)}$

$$\mathbf{u}^{(l)}[0] = 0$$

$$\mathbf{h}^{(l)}[0] = 0$$

$$\mathbf{q}^{(l)}[0] = 0$$

**for**  $t = 1, 2, \dots, T$  **do**

Update  $\mathbf{h}^{(l)}[t]$  based on (7)

Update  $\mathbf{u}^{(l)}[t]$  and  $\mathbf{o}^{(l)}[t]$  based on (1) and (2)

Update  $\mathbf{q}^{(l)}[t]$  based on (5)

**if**  $t \geq t_l$  **then**

Compute  $\mathbf{m}^{(l)}[t]$  based on (9)

Compute  $\Delta \mathbf{W}^{(l)}[t]$  based on (10), (11), and (12)

**end if**

**end for**

$$\mathbf{W}^{(l)} = \mathbf{W}^{(l)} + \eta \sum_{t=t_l}^T \Delta \mathbf{W}^{(l)}[t]$$


---

*1) Memory Requirements:* We begin by discussing memory requirements, focusing on the overhead associated with synaptic updates, excluding the state variables required for SNN inference (e.g., membrane potential). According to [22], the memory requirements for BPTT and S-TLLR can be estimated using the following equations:

$$\text{Mem}_{\text{BPTT}} = T \sum_{l=0}^L n^{(l)}, \quad \text{Mem}_{\text{S-TLLR}} = 2 \sum_{l=0}^L n^{(l)}, \quad (13)$$

where,  $n^{(l)}$  represents the number of neurons in layer  $l$ ,  $L$  is the total number of layers in the model, and  $T$  is the total number of time steps (length of the input sequence). The factor of 2 in (13) arises from the inclusion of both causal and non-causal terms in the computation of eligibility traces when  $\alpha_{\text{pre}}$  and  $\alpha_{\text{post}}$  are nonzero.

For TESS, the memory requirements are determined by analyzing the variables involved in (10) and (11). From (10),  $\mathbf{m}^{(l)}[t]$  depends on the output spikes  $\mathbf{o}^{(l)}[t]$ , which are computed using (9). Since this signal is derived directly from the current output spikes, it does not require additional memory storage and can be computed on the fly. Similarly,  $\Psi(\mathbf{u}^{(l)}[t])$  is a function of the current membrane potential and does not require additional memory, as it can also be computed on the fly. In contrast, the term  $\mathbf{q}^{(l)}[t]$  accounts for the history of pre-synaptic activity and requires memory proportional to the number of input neurons,  $n^{(l-1)}$ . Likewise,  $\mathbf{h}^{(l)}[t]$  represents the history of post-synaptic activity and requires memory proportional to the number of output neurons,  $n^{(l)}$ . By combining these terms, the total memory requirement for TESS can be expressed as:

$$\text{Mem}_{\text{TESS}} = 2 \sum_{l=0}^L n^{(l)}. \quad (14)$$

This demonstrates that TESS achieves memory efficiency comparable to S-TLLR while avoiding the significant overhead associated with the time-dependent storage of BPTT.

2) *Computational Requirements*: Here, we estimate the computational requirements by evaluating the number of MAC operations needed to compute the learning signals. Specifically, we compare the operations required to calculate  $\frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t]}$  for BPTT,  $[\frac{\partial \mathcal{L}}{\partial \mathbf{u}^{(l)}[t]}]_{\text{local}}$  for S-TLLR, and  $\mathbf{m}^{(l)}[t]$  for TESS. For simplicity, we assume a fully connected network and disregard any element-wise operations.

For both BPTT and S-TLLR, the error signals are computed by propagating errors from the last layer to the first. If the final error vector has a dimension of  $n^{(L)}$ , it is propagated to the previous layer using the weight matrix  $\mathbf{W}^{(L)}$ , which has a dimension of  $n^{(L)} \times n^{(L-1)}$ . This matrix-vector multiplication requires  $n^{(L)} \times n^{(L-1)}$  MAC operations. The same process is repeated for all hidden layers. However, while BPTT performs this backpropagation for all time steps  $T$ , S-TLLR computes the learning signal only for the final  $T - t_l$  time steps. Thus, the number of MAC operations can be expressed as:

$$\text{MAC}_{\text{BPTT}} = T \sum_{l=1}^L n^{(l)} \times n^{(l-1)}, \quad (15)$$

$$\text{MAC}_{\text{S-TLLR}} = (T - t_l) \sum_{l=1}^L n^{(l)} \times n^{(l-1)}. \quad (16)$$

In contrast to BPTT and S-TLLR, TESS generates the learning signal locally for each layer, significantly reducing the computational complexity associated with backpropagating errors through layers and time. The TESS learning signal is computed using (9), where the output spikes  $\mathbf{o}^{(l)}[t]$  (of dimension  $n^{(l)}$ ) are projected into a task subspace of dimension  $C$  using a binary fixed matrix  $\mathbf{B}^{(l)}$  of dimension  $C \times n^{(l)}$ . This projection requires  $2 \times C \times n^{(l)}$  MAC operations to compute  $\mathbf{m}^{(l)}$ . Therefore, the total number of MAC operations for a network with  $L$  layers is:

$$\text{MAC}_{\text{TESS}} = (T - t_l) \sum_{l=1}^L 2 \times n^{(l)} \times C \quad (17)$$

where  $C$  represents the number of classes in a classification task or the number of variables to estimate in a regression task. By generating the learning signal locally, TESS achieves a significant reduction in computational complexity compared to BPTT and S-TLLR. Specifically, the reduction factor is approximately  $\frac{n^{(l)}}{C}$ , as  $C \ll n^{(l)}$  in most practical scenarios.

### C. Comparison with other local learning rules

In this subsection, we analyze the time and memory complexity of TESS in comparison to other approaches. For this analysis, we consider a fully connected spiking neural network with  $L$  layers, each containing  $n$  LIF neurons, trained on temporal tasks with  $T$  time steps, and a target space of dimensionality  $C$ .

As discussed in Section II-B, BPTT requires access to the entire history of the network, resulting in a memory

TABLE I  
COMPARISON OF TESS WITH OTHER LEARNING RULES FOR TRAINING SPIKING NEURAL NETWORKS (SNNs) WITH LIF NEURONS. THE PARAMETERS ARE DEFINED AS FOLLOWS:  $L$  REPRESENTS THE NUMBER OF LAYERS,  $n$  IS THE AVERAGE NUMBER OF NEURONS PER LAYER,  $T$  IS THE TOTAL NUMBER OF TIME STEPS, AND  $C$  DENOTES THE NUMBER OF TARGETS.

Method	Memory Complexity	Time Complexity	Temporal Locality	Spatial Locality
BPTT	$TLn$	$TLn^2$	✗	✗
e-prop [9]	$Ln^2$	$Ln^2$	✓	✗
OSTL [10]	$Ln^2$	$Ln^2$	✓	✗
FELL [21]	$Ln^2$	$LCn$	✓	✓
ETLP [19]	$Ln^2$	$LCn$	✓	✓
OSTTP [20]	$Ln^2$	$LCn$	✓	✓
OTTT [23]	$Ln$	$Ln^2$	✓	✗
S-TLLR [22]	$Ln$	$Ln^2$	✓	✗
<b>TESS (Ours)</b>	$Ln$	$LCn$	✓	✓

complexity of  $O(TLn)$ . Similarly, since learning signals are produced by propagating errors through layers for all time steps, the time complexity is  $O(TLn^2)$ . This implies that tasks with greater temporal dependencies significantly increase the cost of BPTT.

To address this dependency on time steps, temporal local methods such as e-prop [9], OSTL [10], OTTT [23], and S-TLLR [22], as well as fully local methods like OSTTP [20], FELL [21] and ETLP [19], leverage eligibility traces. This strategy allows them to maintain a memory requirement independent of time steps. However, methods like e-prop, OSTL, FELL, ETLP, and OSTTP exhibit a memory complexity of  $O(Ln^2)$ , which can become prohibitively expensive for large models. In contrast, methods such as OTTT and S-TLLR achieve a more efficient linear memory complexity of  $O(Ln)$ , making them more scalable. Similarly, TESS has been designed to exhibit linear memory complexity.

Regarding time complexity, methods such as e-prop, OSTL, OTTT, and S-TLLR rely on backpropagation of errors through layers to generate learning signals, incurring a complexity of  $O(Ln^2)$ . In contrast, OSTTP and ETLP use the DRTP mechanism [14] to achieve spatial locality, reducing the time complexity to  $O(LCn)$ , where  $C \ll n$ . Although FELL uses an auxiliary spiking classifier layer that generates error signals by backpropagation through the auxiliary classifier, it still achieves a time complexity of  $O(LCn)$ . TESS follows a similar approach, generating learning signals locally and achieving the same reduced time complexity of  $O(LCn)$ .

Compared to other methods, TESS offers the best combination of memory and time complexity due to its spatial and temporal locality features. Furthermore, TESS achieves this efficiency while delivering performance comparable to methods with higher memory and time requirements, as discussed in the next section. A summary of this comparison is presented in Table I.

TABLE II

ABLATION STUDY ON INCLUDING NON-CAUSAL TERMS ( $\alpha_{\text{post}}$ ) IN THE ELIGIBILITY TRACES DURING LEARNING. ACCURACY (MEAN $\pm$ STD) REPORTED OVER FIVE INDEPENDENT TRIALS

Dataset	$T$	$\alpha_{\text{post}} = -1$	$\alpha_{\text{post}} = 0$	$\alpha_{\text{post}} = +1$
CIFAR10-DVS	10	75.00 $\pm$ 0.69%	<b>75.00 <math>\pm</math> 0.65%</b>	74.36 $\pm$ 0.87%
DVS Gesture	20	98.56 $\pm$ 0.41%	98.33 $\pm$ 0.57%	<b>98.56 <math>\pm</math> 0.31%</b>
CIFAR10	6	89.93 $\pm$ 0.31%	91.99 $\pm$ 0.19%	<b>92.55 <math>\pm</math> 0.16%</b>
CIFAR100	6	62.49 $\pm$ 1.05%	68.19 $\pm$ 0.55%	<b>70.00 <math>\pm</math> 0.34%</b>

#### IV. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of our training algorithm, TESS, on multiple datasets, assessing its ability to achieve competitive accuracy at reduced cost. To do so, we compare TESS with a broad range of non-local to local learning state-of-the-art methods.

##### A. Experimental Setup

This subsection describes the datasets, pre-processing steps, and model architectures used to evaluate our method.

1) *Datasets*: We evaluated TESS using four datasets: CIFAR10 [27], CIFAR100 [27], IBM DVS Gesture [28], and CIFAR10-DVS [29]. The preprocessing steps for each dataset are as follows:

- CIFAR10 and CIFAR100: Images were presented to the SNN models for 6 time steps to simulate a temporal dimension. Data augmentation during training included increasing image size via zero-padding of 4, random cropping to  $32 \times 32$ , application of the cutout technique [30], random horizontal flipping, and normalization.
- CIFAR10-DVS: Events were accumulated in 10 frame events and resized to  $48 \times 48$ . Data augmentation involved random cropping with zero-padding of 4, followed by normalization.
- IBM DVS Gesture: Sequences of varying lengths were split into samples of 1.5 seconds. Events were accumulated into 20 event frames, each representing 75 ms, resized to  $32 \times 32$ , and randomly cropped with zero-padding of 4.

2) *Training details*: All experiments were conducted on a VGG-9 model, in which the ReLU activation functions were replaced with LIF activations. The models were trained for 200 epochs using the Adam optimizer with a learning rate of 0.001. A learning rate scheduler was employed to reduce the learning rate by half if the validation accuracy did not improve for 5 consecutive epochs. The exponential decay factors,  $\lambda_{\text{pre}}$  and  $\lambda_{\text{post}}$ , were set to 0.5 and 0.2, respectively, while  $\alpha_{\text{pre}}$  was fixed at 1. The leaky factor ( $\gamma$ ) and the threshold ( $v_{\text{th}}$ ) of the LIF model were set to 0.5 and 0.6, respectively. Additionally, the secondary activation function was chosen to be a triangular function, defined as  $\Psi(\mathbf{u}) = 0.3 \cdot \max(1 - |\mathbf{u} - v_{\text{th}}|, 0)$ . Finally, weight updates occurred at every time step, i.e.  $t_l = 0$ .

##### B. Results

This subsection presents the results of using TESS across the four datasets, highlighting its performance on image classification tasks and its sensitivity to hyperparameters through ablation studies.

1) *Ablation Studies on Eligibility Traces*: First, we examine the effect of the  $\alpha_{\text{post}}$  parameter to include or exclude the non-causal contribution on the learning process. To assess the influence of this parameter, we set  $\alpha_{\text{post}}$  to values of  $-1$  (negative inclusion),  $+1$  (positive inclusion), and  $0$  (exclusion). We then trained a VGG-9 model on the four datasets described in Section IV-A1, performing five independent trials. The mean and standard deviation of the results are reported in Table II.

From the results, we observed that, with the exception of the CIFAR10-DVS dataset, the positive inclusion of  $\alpha_{\text{post}}$  improves model performance across all other datasets, with gains ranging from 0.23 to 1.81 accuracy points compared to when the parameter is excluded. In contrast, the negative inclusion of  $\alpha_{\text{post}}$  provides a performance improvement only for the IBM DVS Gesture dataset.

Beyond performance, note that the inclusion of the  $\alpha_{\text{post}}$  parameter also impacts the memory usage of TESS. As discussed in Section III-B1, when  $\alpha_{\text{post}} \neq 0$ , additional memory must be allocated for storing the trace of output spikes,  $\mathbf{h}^{(l)}[t]$ . Thus, while the inclusion of  $\alpha_{\text{post}}$  can enhance model performance, it effectively doubles the memory requirements of the algorithm, allowing for a trade-off between performance and memory usage.

2) *Performance on Image Classification*: In this subsection, we compare the performance of TESS with that of previously reported methods, including non-local, partially local, and fully local learning approaches, on the four datasets described in Section IV-A1. The results are presented in Table III.

On CIFAR10-DVS, TESS performs on par with prior methods, including those based on backpropagation and temporally local approaches such as [22], [23], with a maximum accuracy drop of 2.27%, except for [32]. A similar trend is observed on CIFAR10 and CIFAR100, where TESS demonstrates a maximum accuracy drop of approximately 1%. In contrast, on the DVS Gesture dataset, TESS achieves the best performance among all previously reported methods, including non-local approaches, despite being fully local. Notably, TESS outperforms [21], [24], the other fully local methods, by approximately 3 accuracy points. These results highlight the capability of TESS to train models in a fully local manner while achieving performance comparable to or better than non-fully local methods.

It is worth noting that previous works may differ in experimental implementations, introducing variances that are challenging to quantify in the final results. To address this, we established two baselines for each dataset using BPTT and S-TLLR [22], ensuring consistent model implementations, data preprocessing, and hyperparameter settings. Relative to these baselines, TESS outperforms S-TLLR on DVS Gesture, CIFAR10, and CIFAR100 while performing on par with or

TABLE III  
COMPARISON OF PERFORMANCE AND COMPUTATIONAL REQUIREMENTS OF DIFFERENT LOCAL AND NON-LOCAL LEARNING STRATEGIES ON IMAGE CLASSIFICATION TASKS

Method	Model	Local Learning	Time-steps ( $T$ )	Batch Size	Accuracy <sup>1</sup> (mean $\pm$ std)	# MAC <sup>2</sup> ( $\times 10^6$ )	Memory <sup>2</sup> (MB)
CIFAR10-DVS							
BPTT [31]	PLIF (7 layers)	No	20	16	74.8%	-	-
TET [32]	VGG-11	No	10	128	83.17 $\pm$ 0.15%	-	-
DSR [33]	VGG-11	No	10	128	77.27 $\pm$ 0.24%	-	-
OTT <sub>A</sub> [23]	VGG-9	Partial (in time)	10	128	76.27 $\pm$ 0.05%	-	-
BPTT [22]	VGG-9	No	10	48	75.44 $\pm$ 0.76%	-	-
S-TLLR [22]	VGG-9	Partial (in time)	10	48	75.6 $\pm$ 0.10%	-	-
<b>BPTT (baseline)</b>	VGG-9	No	10	64	76.40 $\pm$ 0.66%	13589.59	25.50
<b>S-TLLR (baseline)</b>	VGG-9	Partial (in time)	10	64	75.14 $\pm$ 1.37%	13589.59	5.10
<b>TESS (ours)</b>	VGG-9	Yes	10	64	75.00 $\pm$ 0.65%	22.15	2.55
IBM DVS Gesture							
SLAYER [34] <sup>3</sup>	SNN (8 layers)	No	300	-	93.64 $\pm$ 0.49%	-	-
DECOLLE [24] <sup>4</sup>	SNN (4 layers)	Yes	1800	72	95.54 $\pm$ 0.16%	-	-
OTT <sub>A</sub> [23] <sup>3</sup>	VGG-9	Partial (in time)	20	16	96.88%	-	-
BPTT [22]	VGG-9	No	20	16	95.58 $\pm$ 1.08%	-	-
S-TLLR [22]	VGG-9	Partial (in time)	20	16	97.72 $\pm$ 0.38%	-	-
<b>BPTT (baseline)</b>	VGG-9	No	20	16	97.95 $\pm$ 0.68%	12079.69	22.69
<b>S-TLLR (baseline)</b>	VGG-9	Partial (in time)	20	16	98.48 $\pm$ 0.37%	12079.69	2.26
<b>TESS (Ours)</b>	VGG-9	Yes	20	16	98.56 $\pm$ 0.31%	22.65	2.26
CIFAR10							
Hybrid Training [35]	VGG-11	No	250	128	92.22%	-	-
ELL [21]	SNN (8 layers)	Yes	20	100	89.40 $\pm$ 0.14%	-	-
FELL [21]	SNN (8 layers)	Yes	20	100	88.36 $\pm$ 0.16%	-	-
OTT <sub>A</sub> [23]	VGG-9	Partial (in time)	6	128	93.52 $\pm$ 0.06%	-	-
<b>BPTT (baseline)</b>	VGG-9	No	6	128	92.55 $\pm$ 0.06%	3623.90	6.83
<b>S-TLLR (baseline)</b>	VGG-9	Partial (in time)	6	128	91.88 $\pm$ 0.28%	3623.90	2.27
<b>TESS (ours)</b>	VGG-9	Yes	6	128	92.55 $\pm$ 0.16%	5.48	2.27
CIFAR100							
Hybrid Training [35]	VGG-11	No	250	128	67.87%	-	-
OTT <sub>A</sub> [23]	VGG-9	Partial (in time)	6	128	71.05 $\pm$ 0.04%	-	-
<b>BPTT (baseline)</b>	VGG-9	No	6	128	69.28 $\pm$ 0.37%	3624.18	6.83
<b>S-TLLR (baseline)</b>	VGG-9	Partial (in time)	6	128	68.00 $\pm$ 0.71%	3624.18	2.27
<b>TESS (ours)</b>	VGG-9	Yes	6	128	70.00 $\pm$ 0.34%	17.64	2.27

<sup>1</sup>: Previous studies' accuracy values are provided as reported in their respective original papers.

<sup>2</sup>: # MAC and Memory are estimated for a batch size of 1 with equations presented in Section III-B2.

<sup>3</sup>: These studies used an input resolution of  $128 \times 128$  for the IBM DVS Gesture dataset.

<sup>4</sup>: [24] uses Spike Response Model (SRM) neurons, whereas all other models use LIF neurons.

slightly better than BPTT. The only exception is on CIFAR10-DVS, where TESS lags behind BPTT by 1.4 and S-TLLR by 0.14 accuracy points. Furthermore, TESS achieves these results while significantly reducing the computational cost of generating learning signals, with a reduction in MAC operations of  $205 - 661\times$  thanks to its local learning signal generation. Similarly, TESS reduces memory usage by a factor of  $3 - 10\times$  compared to BPTT.

These findings clearly demonstrate the ability of TESS to train SNN models with accuracy comparable to BPTT while dramatically reducing computational and memory requirements. This makes TESS a highly suitable candidate for enabling learning on low-power devices with constrained resources.

## V. CONCLUSIONS AND PERSPECTIVES

We introduced TESS, a temporally and spatially local learning rule for SNNs, designed to meet the demand for

low-power, scalable training on edge devices. TESS achieves competitive accuracy with BPTT while reducing memory complexity from  $O(TLn)$  to  $O(Ln)$  and time complexity from  $O(TLn^2)$  to  $O(LCn)$ , making it highly efficient for resource-constrained systems. Inspired by biological mechanisms like eligibility traces, STDP, and neural activity synchronization, TESS assigns temporal and spatial credits locally, eliminating the need for global information flow. Experiments demonstrate that TESS has accuracy on par with BPTT on datasets such as CIFAR10, CIFAR100, and IBM DVS Gesture, while losing only 1.4 accuracy points on CIFAR10-DVS. Moreover, it significantly reduces computational costs, with  $205 - 661\times$  fewer MACs and  $3 - 10\times$  lower memory usage. By leveraging its local learning paradigm, TESS offers a scalable, energy-efficient alternative for training SNNs on edge devices, enabling real-time applications with minimal resource demands. Its design is particularly promising for low-power on-device

learning hardware, where spatiotemporal locality is critical for efficiency [4].

## REFERENCES

- [1] Y. Abadade, A. Temouden, H. Bamoumen, N. Benamar, Y. Chtouki, and A. S. Hafid, "A Comprehensive Survey on TinyML," *IEEE Access*, vol. 11, pp. 96 892–96 922, 2023.
- [2] R. Singh and S. S. Gill, "Edge AI: A survey," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71–92, 1 2023.
- [3] V. Tsoukas, A. Gkogkidis, E. Boumpa, and A. Kakarountas, "A Review on the emerging technology of TinyML," *ACM Computing Surveys*, vol. 56, no. 10, 6 2024.
- [4] C. Frenkel and G. Indiveri, "ReckOn: A 28nm Sub-mm<sup>2</sup> Task-Agnostic Spiking Recurrent Neural Network Processor Enabling On-Chip Learning over Second-Long Timescales," in *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, vol. 2022-February, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 468–470.
- [5] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han, "On-Device Training Under 256KB Memory," in *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 6 2022.
- [6] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature*, vol. 575, no. 7784, pp. 607–617, 11 2019.
- [7] D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. Le Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S.-J. Liang, F. Miao, M. Lanza, T. J. Quill, S. T. Keene, A. Salleo, J. Grollier, D. Marković, A. Mizrahi, P. Yao, J. J. Yang, G. Indiveri, J. P. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. H. P. Pernice, H. Bhaskaran, S. Furber, E. Neftci, F. Scherr, W. Maass, S. Ramaswamy, J. Tapson, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T. A. Cleland, C. Posch, S. Liu, G. Panuccio, M. Mahmud, A. N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M. E. Christensen, S. Holm, D. Ielmini, and N. Prysds, "2022 roadmap on neuromorphic computing and engineering," *Neuromorphic Computing and Engineering*, vol. 2, no. 2, p. 022501, 6 2022.
- [8] J. K. Eshraghian, M. Ward, E. O. Neftci, X. Wang, G. Lenz, G. Dwivedi, M. Bennamoun, D. S. Jeong, and W. D. Lu, "Training Spiking Neural Networks Using Lessons From Deep Learning," *Proceedings of the IEEE*, vol. 111, no. 9, pp. 1016–1054, 9 2023.
- [9] G. Bellec, F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, "A solution to the learning dilemma for recurrent networks of spiking neurons," *Nature Communications*, vol. 11, no. 1, 12 2020.
- [10] T. Bohnstingl, S. Wozniak, A. Pantazi, and E. Eleftheriou, "Online Spatio-Temporal Learning in Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [11] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, "Random synaptic feedback weights support error backpropagation for deep learning," *Nature Communications*, vol. 7, 11 2016.
- [12] A. N. Trondheim, "Direct Feedback Alignment Provides Learning in Deep Neural Networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [13] B. Crafton, A. Parihar, E. Gebhardt, and A. Raychowdhury, "Direct feedback alignment with sparse connections for local learning," *Frontiers in Neuroscience*, vol. 13, no. MAY, 2019.
- [14] C. Frenkel, M. Lefebvre, and D. Bol, "Learning Without Feedback: Fixed Random Learning Signals Allow for Feedforward Training of Deep Neural Networks," *Frontiers in Neuroscience*, vol. 15, p. 629892, 2 2021.
- [15] G. Dellaferrera and G. Kreiman, "Error-driven Input Modulation: Solving the Credit Assignment Problem without a Backward Pass," in *Proceedings of the 39th International Conference on Machine Learning*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds. PMLR, 7 2022, pp. 4937–4955.
- [16] G. Hinton, "The Forward-Forward Algorithm: Some Preliminary Investigations," Tech. Rep., 2022. [Online]. Available: <https://www.cs.toronto.edu/~hinton/FFA13.pdf>
- [17] M. P. E. Apolinario, A. Roy, and K. Roy, "LLS: Local Learning Rule for Deep Neural Networks Inspired by Neural Activity Synchronization," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2 2025, pp. 7807–7816. [Online]. Available: <https://ieeexplore.ieee.org/document/10943331/>
- [18] W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, and J. Brea, "Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules," *Frontiers in Neural Circuits*, vol. 12, 7 2018.
- [19] F. M. Quintana, F. Perez-Pefia, P. L. Galindo, E. O. Neftci, E. Chicca, and L. Khacef, "ETLP: event-based three-factor local plasticity for online learning with neuromorphic hardware," *Neuromorphic Computing and Engineering*, vol. 4, no. 3, p. 034006, 8 2024.
- [20] T. Ortner, L. Pes, J. Gentinetta, C. Frenkel, and A. Pantazi, "Online Spatio-Temporal Learning with Target Projection," in *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems*, 4 2023.
- [21] C. Ma, R. Yan, Z. Yu, and Q. Yu, "Deep Spike Learning with Local Classifiers," *IEEE Transactions on Cybernetics*, vol. 53, no. 5, pp. 3363–3375, 5 2023.
- [22] M. P. E. Apolinario and K. Roy, "S-TLLR: STDP-inspired Temporal Local Learning Rule for Spiking Neural Networks," *Transactions on Machine Learning Research*, 1 2025. [Online]. Available: <https://openreview.net/forum?id=CNaiJRcX84>
- [23] M. Xiao, Q. Meng, Z. Zhang, D. He, and Z. Lin, "Online Training Through Time for Spiking Neural Networks," in *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 10 2022.
- [24] J. Kaiser, H. Mostafa, and E. Neftci, "Synaptic Plasticity Dynamics for Deep Continuous Local Learning (DECOLLE)," *Frontiers in Neuroscience*, vol. 14, 5 2020.
- [25] G. Martín-Sánchez, S. Bohté, and S. Otte, "A Taxonomy of Recurrent Learning Rules," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Science and Business Media Deutschland GmbH, 2022, vol. 13529 LNCS, pp. 478–490.
- [26] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 51–63, 11 2019.
- [27] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," 2009.
- [28] A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, and D. Modha, "A Low Power, Fully Event-Based Gesture Recognition System," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2017-January. IEEE, 7 2017, pp. 7388–7397.
- [29] H. Li, H. Liu, X. Ji, G. Li, and L. Shi, "CIFAR10-DVS: An event-stream dataset for object classification," *Frontiers in Neuroscience*, vol. 11, no. MAY, p. 309, 5 2017.
- [30] T. Devries and G. W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout," *arXiv.org*, 2017.
- [31] W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian, "Incorporating Learnable Membrane Time Constant To Enhance Learning of Spiking Neural Networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 2661–2671.
- [32] S. Deng, Y. Li, S. Zhang, and S. Gu, "Temporal Efficient Training of Spiking Neural Network via Gradient Re-weighting," in *International Conference on Learning Representations (ICLR)*, 2022.
- [33] Q. Meng, M. Xiao, S. Yan, Y. Wang, Z. Lin, and Z.-Q. Luo, "Training High-Performance Low-Latency Spiking Neural Networks by Differentiation on Spike Representation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2022, pp. 12 434–12 443.
- [34] S. B. Shrestha and G. Orchard, "SLAYER: Spike Layer Error Reassignment in Time," *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [35] N. Rath, G. Srinivasan, P. Panda, and K. Roy, "Enabling Deep Spiking Neural Networks with Hybrid Conversion and Spike Timing Dependent Backpropagation," in *International Conference on Learning Representations 2020*, 5 2020.